

## AMENDMENTS TO THE CLAIMS

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1-3. *(canceled)*

4. *(currently amended)* The fact extraction tool set of claim ~~3~~7, wherein the attributes include tokenization, text normalization, part of speech tags, sentence boundaries, parse trees, and semantic attribute tagging.

5. *(currently amended)* The fact extraction tool set of claim ~~4~~7, wherein the means for annotating the text comprises a plurality of independent annotators, wherein each of the annotators has at least one specific annotation function, and wherein the fact extraction tool set further comprises user-implemented means for specifying which of the annotators to use and the order of their use.

6. *(canceled)*

7. *(currently amended)* ~~The fact extraction tool set of claim 6, further comprising A~~  
fact extraction tool set for extracting information from a document, wherein the document  
includes text, comprising:

means for breaking the text into tokens;

means for annotating the text with token attributes, constituent attributes, links, and tree-  
based attributes, using XML as a basis for representing the annotated text and for resolving  
conflicting annotation boundaries in the annotated text to produce well-formed XML from the  
results of independent annotators; and

means for extracting facts from the annotated text using text pattern recognition rules,  
wherein each text pattern recognition rule comprises a pattern that describes text of interest, a  
label that names the pattern for testing and debugging purposes, and an action that indicates what  
should be done in response to a matching of the pattern, and wherein the text pattern recognition  
rules use regular expression-based functionality, tree-based functionality, and user-defined  
matching functions.

8. *(currently amended)* The fact extraction tool set of claim ~~3~~7, wherein ~~the means for breaking the text into its base tokens and annotating the base tokens and patterns of base tokens comprises independent annotators, wherein the annotators are of three types comprising:~~

the token attributes, ~~which~~ have a one-per-base-token alignment, where for the attribute type represented, there is an attempt to assign an attribute to each base token;

the constituent attributes are assigned yes-no values, where the entire pattern of each base token is considered to be a single constituent with respect to some annotation value; and

the links, ~~which~~ assign common identifiers to coreferring and other related patterns of base tokens.

9. *(canceled)*

10. *(canceled)*

11. *(currently amended)* The fact extraction tool set of claim ~~4~~12, wherein the means for identifying and extracting potentially interesting pieces of information ~~comprises means for~~ performs the further function of recognizing both true left and right constituent attributes and non-contiguous constituent attributes.

12. ~~(currently amended)~~ ~~The fact extraction tool set of claim 10;~~ A fact extraction tool set for extracting information from a document, wherein the document includes text, comprising:  
means for breaking the text into tokens;  
means for annotating the text with token attributes, constituent attributes, links, and tree-based attributes, using XML as a basis for representing the annotated text;  
means for associating all annotations assigned to a particular piece of text with the base tokens for that text to generate aligned annotations; and  
~~wherein the~~ means for identifying and extracting potentially interesting pieces of information ~~comprises at least one text pattern recognition rule in the aligned annotations by finding patterns in the attributes of the annotated text using text pattern recognition rules written in a rule-based information extraction language, wherein each text pattern recognition rule comprises a pattern that describes text of interest, a label that names the pattern for testing and debugging purposes, and an action that indicates what should be done in response to a matching of the pattern, and wherein the text pattern recognition rules use regular expression-based functionality, tree-based functionality, and user-defined matching functions, and wherein the at least one each~~ text pattern recognition rule queries for at least one of literal text, attributes, and relationships found in the aligned annotations to define the facts to be extracted.

13-14. *(canceled)*

15. *(currently amended)* The fact extraction tool set of claim 12, wherein the ~~means for identifying and extracting potentially interesting pieces of information further comprises at least one auxiliary definition statement~~ user-defined matching functions are used to name and define a fragment of a pattern.

16. *(currently amended)* A rule-based information extraction language for use in identifying and extracting potentially interesting pieces of information in aligned annotations in a text, ~~the language comprising at least one a plurality of text pattern recognition rules~~ rules that queries ~~query~~ for at least one of literal text, attributes, and relationships found in the aligned annotations to define the facts to be extracted, using regular expression functionality, tree-based functionality, and auxiliary definitions in any combination wherein each of the text pattern recognition rule comprises:

a pattern that describes text of interest;

a label that names the pattern for testing and debugging purposes; and

an action that indicates what should be done in response to a matching of the pattern; and

wherein the text pattern recognition rules use regular expression-based functionality, tree-based functionality, and user-defined matching functions.

17-18. *(canceled)*

19. *(currently amended)* The language of claim 16, ~~further comprising at least one auxiliary definition statement~~ wherein the user-defined matching functions are used to name and define a fragment of a pattern.

20. (*currently amended*) A text annotation tool comprising:

means for breaking text into its base tokens;

~~means a plurality of independent annotators for annotating the base tokens and patterns of~~  
~~base tokens with the text with token attributes, constituent attributes, links, and tree-based~~  
~~attributes, using XML as a basis for representing the annotated text, wherein each of the~~  
~~annotators has at least one specific annotation function;~~

~~user-implemented means for specifying which of the annotators to use and the order of~~  
~~their use; and~~

means for associating all annotations assigned to a particular piece of text with the base  
tokens for that particular piece of text to generate aligned annotations; and

means for resolving conflicting annotation boundaries in the annotated text to produce  
well-formed XML.

21. (*previously presented*) The text annotation tool of claim 20, wherein the attributes  
include tokenization, text normalization, part of speech tags, sentence boundaries, parse trees,  
and semantic attribute tagging.

22-24. (*canceled*)

25. *(currently amended)* The text annotation tool of claim 20, wherein ~~the means for~~  
~~annotating the base tokens and patterns of base tokens comprises independent annotators,~~  
~~wherein the annotators are of three types comprising:~~

the token attributes, ~~which~~ have a one-per-base-token alignment, where for the  
attribute type represented, there is an attempt to assign an attribute to each base token;

the constituent attributes are assigned yes-no values, where the entire pattern of  
each base token is considered to be a single constituent with respect to some annotation value;  
and

where the links, ~~which~~ assign common identifiers to coreferring and other related  
patterns of base tokens.

26-40. *(canceled)*

41. *(currently amended)* A method of extracting information from a document, wherein the document includes text, comprising the steps of:

breaking the text into base tokens;

annotating the text with regular, ~~expression-based~~ attributes and with tree-based attributes; and token attributes, constituent attributes, links, and tree-based attributes, using XML as a basis for representing the annotated text;

resolving conflicting annotation boundaries in the annotated text to produce well-formed XML; and

extracting facts from the annotated text using text pattern recognition rules written in rule-based information extraction language, using regular, ~~expression-based~~ functionality, tree-based functionality, and auxiliary definitions in any combination wherein each text pattern recognition rule comprises a pattern that describes text of interest, a label that names the pattern for testing and debugging purposes, and an action that indicates what should be done in response to a matching of the pattern, and wherein the text pattern recognition rules use regular expression-based functionality, tree-based functionality, and user-defined matching functions.

42. *(canceled)*

43. *(currently amended)* The method of claim 41, wherein ~~the parsing of the text comprises breaking it into its base tokens and annotating the base tokens and patterns of base tokens with a number of~~ in the annotating step, the attributes include orthographic, syntactic, semantic, pragmatic and dictionary-based attributes.

44. (canceled)

45. (currently amended) The method of claim 41, wherein the ~~parsing of the text~~  
annotating step is carried out by a plurality of independent annotators, wherein each of the  
annotators has at least one specific annotation function, and wherein the method further  
comprises the step of allowing a user to specify which of the annotators to use and the order of  
their use.

46-47. (canceled)

48. (currently amended) The method of claim ~~[[43]]~~41, wherein ~~the step of breaking~~  
~~the text into its base tokens and annotating the base tokens and patterns of base tokens is carried~~  
~~out using independent annotators, wherein the annotators are of three types comprising:~~

the token attributes, ~~which~~ have a one-per-base-token alignment, where for the  
attribute type represented, there is an attempt to assign an attribute to each base token;

the constituent attributes are assigned yes-no values, where the entire pattern of  
each base token is considered to be a single constituent with respect to some annotation value;  
and

the links, ~~which~~ assign common identifiers to coreferring and other related  
patterns of base tokens.

49. *(currently amended)* The method of claim ~~[[43]]~~41, wherein the ~~step of~~ step a text further comprises the step of includes associating all annotations assigned to a particular piece of text with the base tokens for that text to generate aligned annotations.

50. *(canceled)*

51. *(currently amended)* The method of claim ~~[[50]]~~41, wherein the ~~step of~~ identifying and extracting potentially interesting pieces of information comprises recognizing text pattern recognition rules have the ability to recognize both true left and right constituent attributes and non-contiguous constituent attributes.

52. *(currently amended)* The method of claim ~~[[50]]~~41, wherein the ~~patterns are found using at least one text pattern recognition rule written in a rule-based information extraction language, wherein the at least one text pattern recognition rule queries rules query for~~ at least one of literal text, attributes, and relationships found in the aligned annotations to define the facts to be extracted.

53-55. *(canceled)*

56. *(new)* The fact extraction tool set of claim 7, wherein the pattern recognition rules query for at least one of literal text, attributes, and relationships found in the annotated text to define facts to be extracted.

57. (new) The fact extraction tool set of claim 7, wherein the means for annotating the text represents the annotated text as a single view of the document expressed as inline XML.

58. (new) The fact extraction tool set of claim 12, wherein the means for annotating the text represents the annotated text as a single view of the document expressed as inline XML.

59. (new) The method of claim 41, wherein in the annotating step, the annotated text is represented as a single view of the document expressed as inline XML.

60. (new) The fact extraction tool set of claim 7, wherein the means for extracting uses XPath for traversing XML-based tree representations in the annotated text.

61. (new) The fact extraction tool set of claim 12, wherein the means for extracting uses XPath for traversing XML-based tree representations in the annotated text.

62. (new) The method of claim 41, wherein in the extracting step, XPath is used for traversing XML-based tree representations in the annotated text.